Measuring corporate social responsibility and impact: enhancing quantitative research design and methods in business and society research

An issue of Business & Society does not pass where there is not at least one, if not more, papers on corporate social responsibility (CSR) or corporate social performance (CSP). In fact, the field is obsessed with issues of performance and impact. Does CSR have an impact on financial performance? Or, perhaps more importantly, does CSR have an impact on social welfare or firm environmental performance? These questions go to the heart of our field and are related to its legitimacy as an area of intellectual inquiry. Still the answers to these questions are not clear and the results of research continue to be mixed. These motley results are no doubt due to the many methodological issues that such research faces. Still, progress is being made – slowly. As editors, the main reasons we desk reject quantitative papers are due to inattention to issues such as variable measurement, endogeneity, survey biases (Fowler, 2014) and other methodological issues. As we receive manuscripts from around the world, we see that many studies do not incorporate state-of-the-art measurement and research design that would allow for the more rapid accumulation of knowledge. So in this *Editors' Insight*, we want to draw your attention to ways we, as business and society researchers, can improve our quantitative methods for evaluating the impact of CSR engagement and related initiatives of the firm not only on financial performance, but also on social wellbeing.

Current research measures and designs

The last twenty years have seen the growth of a rich conversation regarding the measures used to measure CSR performance. The social ratings provided by Kinder, Lydenberg, and Domini & Co. (KLD) have constituted the great workhorse of CSR research. Other agencies have now

produced similar kinds of ratings such as Vigeo and Thomson Reuters Asset4, which are now beginning to enter CSR research (Liang & Renneboog, 2017). These ratings have been examined from every direction with some studies finding them to have construct validity (Sharfman, 1996; Hart & Sharfman, 2015), while others have uncovered significant problems. For example, KLD consists of a number of dimensions with respect to both strengths and concerns, which some researchers aggregate into a single construct (El Ghoul et al., 2011), while others maintain separate constructs (Mattingly & Berman, 2006; Strike, Gao, & Bansal, 2006). There is also criticism relating to the aggregation of the different issue areas (community spending, diversity, governance, etc.) into a single construct of corporate social performance (Rowley & Berman, 2000; Capelle-Blancard & Petit, 2017). Fortunately, recent advances, such as those related to item response theory and Bayesian methods, open up new avenues for significantly improving our measures by explicitly modelling firms' and individuals' observable actions such as voluntary corporate policies or actions in order to measure unobserved, latent characteristics such as norms and values (akin to implicit CSR discussed by Matten and Moon (2008)) (Carroll, Primo, & Richter, 2016).

Of even greater concern than the specific indicators used to measure different facets of the participation in and outcomes of a wide variety of CSR initiatives is the research design used to evaluate the impact of corporate initiatives for either social performance or corporate financial performance (CFP). Like research studying the impact of any initiative involving a strategic decision, CSP-CFP research is also plagued with endogeneity (Benabou & Tirole, 2010), which makes causal inference almost impossible. Endogeneity in CSP-CFP research occurs because the decision to engage in CSR is correlated with the error term. In other words, there are unobserved factors that may be driving both the decision and subsequent performance, whether financial, social, or environmental. Endogeneity may be due to omitted variables, measurement error, or

reverse causality. The problem is that endogeneity may result in inconsistently estimated regression coefficients, thus leading to incorrect conclusions about their significance (Hamilton & Nickerson, 2003). A quick review of research regarding the CSP-CFP relationship reveals that the issue of endogeneity was rarely even discussed in CSR research before 2008. It might not be too outrageous to suggest that most CSP-CFP research before 2008 should be held suspect for not treating this issue carefully. In a recent replication of Waddock and Graves (1997), Zhao and Murrell (2016) correct for some of the problems we have identified in this essay and find that the relationship between CSR engagement and FP disappears. In another recent study that corrects for endogeneity, Garcia-Castro, Ariño, and Canela (2010) find that the CSP-CFP relationship disappears. So cause for alarm is warranted. In fact, a company's social engagement is likely endogenous with respect to many activities that business and society researchers are seeking to understand. In this Editors' Insights essay, we will examine some of the most important problems of research design in business and society and review a few of the methods available to remedy these flaws.

Departures from experimental design in business and society research

The classical laboratory experiment provides a simple way of understanding the approaches to solving the many problems in current business and society research. The classical experiment consists of a treatment group, an equivalent control group, observations before and after the treatment, and random assignment of the subjects into the treatment and control groups (Campbell & Stanley, 1971). The challenge in establishing causality is determining what would have happened to the treatment group if it had not received the treatment. In terms of CSR-FP research, for example, the challenge is what would have happened to financial performance in the absence of CSR engagement by firms that did engage in CSR – a mindbending Zen koan, commonly

called a counterfactual. The classical experiment resolves this problem by creating an equivalent control group that is identical to the treatment group in all respects, except the presence of the treatment. This equivalence is achieved by random assignment of the subjects to the treatment or control groups. The difference in the outcome after the application of a treatment is the treatment effect or the impact we are seeking to measure. The key is randomization. Although regarded as a pipe dream by many management researchers, major headway has been made in making randomized experiments feasible at macro levels of analysis. For example, Duflo, Glennerster, and Kremer (2007) provide guidance as to how to create randomization in development economics. In research where performance is the dependent variable, Chatterji, et al. (2016) provide excellent advice for researchers who desire to develop randomized field experiments. In our field, Kistruck, et al. (2013) apply these methods to study agency costs in base-of-the-pyramid markets.

Alas, the conditions for an ideal experiment are not always available. It may not be possible to obtain data before and after the treatment. In other cases, obtaining a control group or random assignment to a control group is not feasible. In such cases, researchers can attempt to simulate experimental conditions.

Let's take the example of the impact of an environmental program on environmental emissions (Henriques, Husted, & Montiel, 2013). It is highly unlikely that a researcher will find conditions where firms will randomly adopt a particular environmental program. A comparison of firms that adopt the program with firms that do not adopt it may reveal that firms that adopt the program may actually have higher emissions than firms that do not adopt the program! Should we conclude that the environmental program has failed? Not just yet. Although we have a control group in terms of firms that do not adopt the program, the difference in the observed emissions of

the control group and the treatment group does not capture the impact of the program. We need to look at emissions before the implementation of the environmental program.

An alternative way to determine the impact might be to look at the change in emissions in the treatment group before and after implementation. Although that might partially capture the impact of the environmental program on the treatment group, there may have been all sorts of external factors like new regulation or an economic slowdown that may account for the reduction in emissions.

The next step is to compare the treatment group and the control group before and after the implementation of the environmental group by treatment group. The difference in the before and after outcomes for the treatment group may in fact be greater than the difference in the before and after outcomes for the control group. However, if assignment to these groups is not randomized, then the two groups may not be entirely comparable and the difference in the results may be due to differences in the two groups, rather than to the implementation of the environmental program itself. In fact, the literature finds just this result. Often more heavily polluting firms are more likely to adopt voluntary environmental programs, so it is not clear whether the reduction in emissions is due to the program itself or due to the fact that the adopters were worse off to begin with and improvement of any sort was easier for any reason. This is called the sample selection problem.

Sample selection bias occurs because firms will choose to participate in a given program based on their performance. So the decision to participate is itself endogenous. In a true experiment, one could simply compare the outcomes of the comparison groups. However, to the extent that the conditions depart from these ideal ones, the evaluation of impact becomes more complex.

Strategies for simulating experimental conditions in business and society research

Given these common problems in research design, we outline a few of the strategies that have begun to revolutionize causal inference in other fields and need to take hold in business and society research. The first issue is related to finding an equivalent control group. One method that is being used with greater frequency is propensity score matching (PSM) (Dehejia & Wahba, 2002), which allows for matching on observed characteristics. Specifically, PSM matches firms based on the probability that a firm will adopt a program given those observed characteristics. In the case of environmental emissions, firms could be matched statistically based on size and industry. This predicted probability of adopting or participating based on the observed characteristics is called the propensity score. PSM assumes that selection bias is related to these observed characteristics and sues the propensity score to create comparable treatment and control groups. Deng et al. (2013) provide a good example of the use of propensity score matching in CSR research.

The limitation of PSM is that it assumes that the treatment and control groups do not differ in unobserved characteristics. In contrast, the double differences approach allows for selection bias in unobserved characteristics, but assumes that these characteristics do not change over time (Khandker, Koolwal, & Samad, 2009). Similar to the ideal experiment, the outcomes are observed both before and after the treatment (the first difference) and with and without the treatment (the second difference). The double differences method is used when randomization is not available, but matching occurs by dropping comparable cases that have propensity scores outside the range of propensity scores of the group the adopts the CSR initiative – in this case an environmental program.

More recently, researchers have begun using quasi-natural experiments to address endogeneity. So in a model where you are looking at whether CSR affects financial performance (i.e., does X cause Y where X is CSR and Y is financial performance), you need an experimental

setting in which CSR (X) varies exogenously. Flammer (2015) and Flammer & Bansal (2017) use such a method design (regression discontinuity design) to compare financial performance below and above a threshold – namely CSR shareholder proposals that pass or fail by a small margin of votes. Such close call CSR shareholder proposals are essentially random outcomes thereby providing the researcher with a randomized assignment of CSR to companies (Flammer, 2015).

Finally, if the unobserved characteristics change over time, then an instrumental variables approach is called for. An instrumental variable (instrument) is one that is related to the treatment (adoption of a program or practice), but not related to the outcome (in our case environmental performance or financial performance). The instrumental variables are used to predict the adoption of the initiative and then the predicted values of adoption or participation is used to determine how adoption relates to changes in the outcome (emissions). The main problem of an instrumental variables approach is that finding good instruments is a significant challenge. Nevertheless, it can be done. Using a related procedure, Shahzad and Sharfman (2017) reexamine the CSR-FP relationship with a four-year panel and take into account firms with KLD scores and those firms without KLD scores. They find a positive relationship between CSR engagement and financial performance.

Conclusion

So where to go from here? First, if you are not familiar with any of these methods or concerns, there are some excellent resources available from neighboring disciplines such as public policy and economics. Martin Ravallion (2001) provides an entertaining introduction to the world of impact evaluation. Closer to home, Hiscox, Schwartz, and Toffel (2009) discuss issues of impact or performance evaluation in the context of SA8000 certification. Their chapter provides another accessible introduction to these methods. Finally, the World Bank has published an excellent handbook on impact evaluation (Khandker, Koolwal, & Samad, 2009), which explains the models in

greater detail, but is still user friendly. Next, we need to make these improvements in our research - immediately.

Now it is important to emphasize that these methods are relevant to many areas of interest for business and society research. Stakeholder theory is often concerned with the impacts of stakeholder management initiatives (Harrison, Bosse, & Phillips, 2010). As mentioned earlier, these approaches are relevant to performance in its broadest sense, not just financial performance, or even the typical triple-bottom-line measures of performance. These methods can also be applied to different levels of analysis. For example, Flammer and Kacperczyk (2016) used a natural experiment design and the difference-in-differences methodology to examine the causal impact of stakeholder orientation on innovation. To deal with the endogeneity between stakeholder orientation and innovation, they exploit the state-level constituency statutes which allowed corporate directors to consider stakeholders' interests – in other words, they have a treatment group composed of the states that adopted the statutes and a control group with the states that had not. Moreover, these methods can be applied just as easily to the individual, state or country levels of analysis.

Undoubtedly quantitative research in business and society has reached its most exciting period. Although the challenges are greater than ever before, including obtaining longitudinal data and using accurate measures and appropriate research designs, the rewards are enormous. Given methodological flaws, we cannot say anything with much confidence about the CSP-CFP relationship or even the CSR-CSP relationship in studies published before 2008! In addition, as Rost & Ehrmann (2017) explain, serious problems of reporting bias in favor of positive outcomes may have led to incorrect conclusions in the literature. Unfortunately, the field as a whole has been somewhat slow to take into account the concerns outlined in this essay. Consequently, we look forward to receiving empirical research that carefully considers these issues and incorporates

appropriate measures and methods to address them. We would also welcome well-crafted replications of prior studies that address these concerns. Such replications would not approach the problem piecemeal, but rather use the whole range of state-of-the art measures and methods that can be used to improve our understanding of CSR engagement and its impacts.

So in line with the theme of this Editors' Insights, we have grouped papers that sought to improve our methods for evaluating the financial, social, and environmental impacts of CSR engagement and related initiatives of the firm. The issue begins with a review paper by Mattingly (2017) of research using KLD data. Methodological issues relating to reporting biases and sample-selection biases in CSR research are taken up by Rost & Ehrmann (2017) and Shahzad & Sharfman (2017) respectively. The issue concludes with a research note on the weighting of CSR dimensions by Capelle-Blancard & Petit (2017). So will we start solving the problems we face in CSR measurement and impact? Stayed tuned to find out!

Andrew Crane
School of Management, University of Bath, United Kingdom
Irene Henriques
Schulich School of Business, York University, Toronto, Ontario, Canada
Bryan W. Husted
EGADE Business School, Tecnológico de Monterrey, Monterrey, Nuevo León, Mexico
Dirk Matten
Schulich School of Business, York University, Toronto, Ontario, Canada

References

Bénabou, R., & Tirole, J. (2010). Individual and corporate social responsibility. *Economica*, *77*(305), 1-19.

Campbell, D. T. & Stanley, J. C. 1971. Experimental and quasi-experimental designs for research. Chicago, Il: Rand McNally & Company.

Capelle-Blancard, G., & Petit, A. (2017). The Weighting of CSR Dimensions: One Size Does Not Fit All. *Business & Society*. DOI: 10.1177/0007650315620118.

Carroll, R. J., Primo, D. M., & Richter, B. K. (2016). Using item response theory to improve measurement in strategic management research: An application to corporate social responsibility. *Strategic Management Journal*, *37*(1), 66-85.

Chatterji, A. K., Levine, D. I., & Toffel, M. W. (2009). How well do social ratings actually measure corporate social responsibility? *Journal of Economics & Management Strategy*, *18*(1), 125-169.

Chatterji, A. K., Findley, M., Jensen, N. M., Meier, S., & Nielson, D. (2016). Field experiments in strategy research. *Strategic Management Journal*, *37*(1), 116-132.

Dehejia, R. H., & Wahba, S. (2002). Propensity score-matching methods for nonexperimental causal studies. *Review of Economics and statistics*, *84*(1), 151-161.

Deng, X., Kang, J. K., & Low, B. S. (2013). Corporate social responsibility and stakeholder value maximization: Evidence from mergers. *Journal of Financial Economics*, *110*(1), 87-109.

Duflo, E., Glennerster, R., & Kremer, M. (2007). Using randomization in development economics research: A toolkit. *Handbook of development economics*, *4*, 3895-3962.

El Ghoul S, Guedhami O, Kwok C, Mishra D. 2011. Does corporate social responsibility affect the cost of capital. *Journal of Banking and Finance* **35**: 2388-2406.

Flammer, C. (2015). Does Corporate Social Responsibility Lead to Superior Financial Performance? A Regression Discontinuity Approach. *Management Science*, 61(11), 2549–2568.
Flammer, C., & Bansal, P. (2017). Does a long-term orientation create value? Evidence from a regression discontinuity. *Strategic Management Journal*. Retrieved from http://doi.wiley.com/10.1002/smj.2629

Flammer, C., & Kacperczyk, A. (2016). The Impact of Stakeholder Orientation on Innovation: Evidence from a Natural Experiment. Management Science, 62(7), 1982–2001.

Fowler, F. J. (2014). Survey research methods. 5th Edition, London, U.K.: Sage Publications.

Garcia-Castro, R., Ariño, M. A., & Canela, M. A. (2010). Does social performance really lead to financial performance? Accounting for endogeneity. *Journal of Business Ethics*, *92*(1), 107-126.

Hamilton, B. H., & Nickerson, J. A. (2003). Correcting for endogeneity in strategic management research. *Strategic organization*, *1*(1), 51-78.

Hart, T. A., & Sharfman, M. (2015). Assessing the concurrent validity of the revised Kinder, Lydenberg, and Domini corporate social performance indicators. *Business & Society*, *54*(5), 575-598.

Harrison, J. S., Bosse, D. A., & Phillips, R. A. (2010). Managing for stakeholders, stakeholder utility functions, and competitive advantage. *Strategic Management Journal*, *31*(1), 58-74.

Hiscox MJ, Schwartz C, & Toffel M. 2009. Evaluating the impact of SA8000 certification. In Leipziger, D. (Ed.) SA8000 The First Decade: Implementation, Influence, and Impact. London: Greenleaf, pp. 1-11.

Henriques, I., Husted, B. W., & Montiel, I. (2013). Spillover effects of voluntary environmental programs on greenhouse gas emissions: Lessons from Mexico. *Journal of Policy Analysis and Management*, 32(2), 296-322.

Ioannou, I., & Serafeim, G. (2012). What drives corporate social performance? The role of nation-level institutions. *Journal of International Business Studies*, *43*(9), 834-864.

Khandker, S. R., Koolwal, G. B., & Samad, H. A. (2009). *Handbook on impact evaluation: quantitative methods and practices*. World Bank Publications.

Kistruck, G. M., Sutter, C. J., Lount, R. B., & Smith, B. R. (2013). Mitigating principal-agent problems in base-of-the-pyramid markets: An identity spillover perspective. *Academy of Management Journal*, *56*(3), 659-682.

Liang, H., & Renneboog, L. (Forthcoming, 2016). On the foundations of corporate social responsibility. *The Journal of Finance*.

Matten, D. and J. Moon (2008): 'Implicit' and 'explicit' CSR: A conceptual framework for a comparative understanding of corporate social responsibility. *Academy of Management Review* , 33 (2), 404-424.

Mattingly, J. E. (2017). Corporate social performance: A review of empirical research examining the corporation-society relationship using Kinder, Lydenberg, Domini social ratings data. *Business & Society*, DOI: 10.1177/0007650315585761.

Mattingly, J. E., & Berman, S. L. (2006). Measurement of corporate social action discovering taxonomy in the Kinder Lydenburg Domini ratings data. *Business & Society*, *45*(1), 20-46.

Ravallion, M. (2001). The mystery of the vanishing benefits: An introduction to impact evaluation. *the world bank economic review*, *15*(1), 115-140.

Rost, K., & Ehrmann, T. (2017). Reporting Biases in Empirical Management Research: The Example of Win-Win Corporate Social Responsibility. Business and Society, DOI: 10.1177/0007650315572858.

Rowley, T., & Berman, S. (2000). A brand new brand of corporate social performance. *Business & society*, *39*(4), 397-418.

Sharfman, M. (1996). The construct validity of the Kinder, Lydenberg & Domini social performance ratings data. *Journal of Business Ethics*, *15*(3), 287-296.

Shahzad, A. M., & Sharfman, M. P. (2017). Corporate social performance and financial performance sample-selection issues. *Business & Society*, DOI: 10.1177/0007650315590399.

Strike V, Gao J, & Bansal P. 2006. Being good while being bad: social responsibility and the international diversification of US firms. *Journal of International Business Studies* 37(6): 850-862.

Waddock, S. A., & Graves, S. B. (1997). The corporate social performance-financial performance link. *Strategic Management Journal*, 18(4). 303-319.

Zhao, X., & Murrell, A. J. (2016). Revisiting the corporate social performance-financial performance link: A replication of Waddock and Graves. *Strategic Management Journal*, *37*(11), 2378-2388.